



# La presse écrite sur le web et ses lecteurs comme limites aux techniques de personnalisation de l'information

Laurent Collet

## ► To cite this version:

Laurent Collet. La presse écrite sur le web et ses lecteurs comme limites aux techniques de personnalisation de l'information. Revue des Sciences et Technologies de l'Information - Série Document Numérique, Lavoisier, 2011, 14/3 - 2011, pp.81-102. <10.3166/dn.14.3.81-102>. <halshs-00731075>

**HAL Id: halshs-00731075**

**<https://halshs.archives-ouvertes.fr/halshs-00731075>**

Submitted on 11 Sep 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

---

# La presse écrite sur le web et ses lecteurs comme limites de la personnalisation de l'information

The press on the web, its readers and the limits to personalized information

Laurent COLLET

MCF Sciences de l'information et de la communication / Laboratoire Pacte

Université du Sud - Toulon - Var

Avenue de l'Université - BP20132

83957 LA GARDE CEDEX – France

@ : laurent.collet@univ-tln.fr

---

*RESUME* La personnalisation d'informations est une tendance devenue importante du web. Elle suppose la construction de représentations des lecteurs et/ou des comportements de lecture, construits sur la base de clics, que les serveurs arrivent à capter : données comportementales, contextuelles (sur quoi le clic a porté) et déclaratives (qui est derrière le clic). Cet article propose, à partir de données recueillies dans le cadre d'une recherche/action, d'interroger les enjeux de cette personnalisation pour les entreprises de presse écrite sur le web, et surtout les limites en termes d'usages et de stratégie économique.

*ABSTRACT* : Personalized information is a growing trend on the web. It relies on representations of the readers and/or their reading habits based on their clicks which computers can perceive and interpret : behavioural, contextual (what the person clicked) and declarative (?) data (who clicked?). This paper; on the basis of data collected as part of a research project, offers to question what is at stake in personalized information, especially for the media on the internet and, above all, what are its limits as far as habits and economic strategies are concerned.

*MOTS-CLÉS* : personnalisation, rationalisation, taylorisation, marchandisation, industrialisation, presse écrite, usage, approche sémiotique

*KEYWORDS*: personalization, rationalization, taylorization, merchandization, industrialization, written press, use, semiotic approach.

---

## 1. Introduction

Vannevar Bush, auteur de l'article « As we may think », qui pose les bases d'un dispositif de personnalisation de l'accès à l'information (MEMEX), visait l'amélioration de la productivité informationnelle des militaires, et plus généralement des êtres humains (BUSH, 1945). Son travail pour l'armée a jeté les bases d'une certaine forme de personnalisation du rapport à l'information, limitée à son archivage. Des années plus tard, Théodor Nelson, via son projet Xanadu, élaborera le terme d'hypertexte pour signifier des liens personnels entre informations ne se limitant pas à lier des citations aux textes originaux ou à ajouter des commentaires. Il propose également de lier des concepts entre eux. Ces liens peuvent être l'œuvre d'auteurs ou de lecteurs rendant par la même floue la distinction entre les deux. Ce même Nelson inspirera Tim Berners-Lee, chercheur au CERN de Lausanne, le créateur du langage HTML et du web. Dès ses fondements, le web a donc à voir avec la personnalisation de l'information.

Des décennies après Bush, les acteurs du web réfléchissent toujours à cette question mais contrairement aux premières pages HTML, qui favorisait de multiples parcours au sein de contenus statiques, la voie aujourd'hui empruntée est double.

La première passe par le développement du web sémantique, dont un des objectifs est de permettre l'interrogation des moteurs de recherche en langage naturel sans se soucier 1) d'avoir à formuler des requêtes sous forme de mots clefs, 2) à devoir consulter et sélectionner les informations recueillies, voire à les comparer pour obtenir une réponse<sup>1</sup>. Cette sémantisation demande d'ajouter des métadonnées aux contenus affichés et visibles par les internautes sur les pages web. Ces données sont destinées aux algorithmes des robots d'indexation des moteurs de recherche. Elle vise donc à faciliter la recherche d'informations par les humains et prolonge les désirs de BUSH et NELSON d'améliorer leur productivité informationnelle (BERNERS-LEE, HENDLER, LASSILA, 2001). Les enjeux marchands, bien que jamais totalement absents, ne sont pas le moteur du web sémantique, au contraire de la seconde voie, qui émerge directement des acteurs à capitaux privés du web.

---

<sup>1</sup> Les promoteurs du web sémantique citent plusieurs exemples pour clarifier ce concept, donc celui de la recherche d'un film dans une salle de cinéma dans une ville. Plutôt que d'aller consulter, sur le web, le programme de toutes les salles de cinéma de sa ville, l'internaute n'aurait qu'à demander à un moteur de recherche « quelle salle de telle ville programme tel film ? ». Le moteur de recherche, grâce aux métadonnées, n'aurait plus qu'à consulter tous les programmes diffusés sur le web pour une ville donnée et de construire une réponse adhoc. Il pourrait proposer des réponses si au lieu d'un titre de film, il est interrogé sur un réalisateur, à condition que les titres diffusés aient la métadonnée réalisateur.

Cette seconde voie est celle de l'interprétation des comportements des internautes, sur la base d'informations, que les serveurs arrivent à capter<sup>2</sup>. Mais contrairement au MEMEX, projet visant l'amélioration des prothèses cognitives au service des êtres humains, les algorithmes de personnalisation actuels visent la taylorisation de la production (conception et diffusion) d'informations pour améliorer leur marchandisation. L'enjeu est industriel car la rationalisation en cours n'est pas la recherche d'une rationalité issue de l'ère des Lumières et ne porte pas sur la formation d'un espace public gouverné par la Raison (HABERMAS, 1962). Elle est plutôt liée aux nouveaux modes de rémunération des régies publicitaires : la rémunération au clic, voire à la transaction engendrée, et non plus au simple affichage de l'information (OUAKRAT et ali., 2010 ; CARON, FRANCONY, QUINETTE, 2011). C'est pourquoi les acteurs du web cherchent à récolter toute une série de données afin de qualifier les adresses électroniques, ou les profils des abonnés, en leur possession :

- les informations déclarées par les utilisateurs d'un service en ligne (âge, sexe, ...). Les jeux concours en ligne sont un bon moyen pour obtenir ces données accolées à une adresse électronique. Pour autant, rien ne garantit que l'internaute fournisse les vraies données le concernant. D'ailleurs, les acteurs savent que beaucoup d'internautes possèdent des adresses électroniques uniquement pour jouer sur le web ; données qui permettent de « mentir » sur son sexe, son âge, etc., et surtout d'être supprimées en cas de surcharge informationnelle.
- les données contextuelles comme les contenus lus par les utilisateurs. C'est le principe de base des publicités contextuelles proposées par la régie publicitaire de Google sous la formule Adsense. Ce principe est aujourd'hui largement répandu et appliqué par toutes les régies publicitaires du web. C'est à partir des seules données contextuelles que la plupart des régies gèrent la personnalisation des annonces commerciales.
- l'analyse comportementale des clics et parcours de lecture, grâce à des outils d'analyse permettant d'analyser les visites. Google essaie de se positionner sur ce marché à travers son logiciel « Google analytics ». Facebook n'est pas en reste sur sa plate-forme de réseau social et sa technologie Social Ads : un bandeau publicitaire vertical, à droite de la page, composé de trois publicités différentes en fonction du profil (l'âge, le lieu de résidence et la profession, éventuellement sa situation matrimoniale) mais aussi de ses comportements (ses hobbies, ses goûts musicaux et cinématographiques, sa religion ou encore ses opinions politiques).

---

<sup>2</sup> Ainsi, les réseaux sociaux (Facebook, ...), les moteurs de recherche (Google et son offre Adsense), les infomédiaires (Wikio, ...), les éditeurs traditionnels de la presse écrite (l'offre Desk du Monde) élaborent, chacun à leur manière, des algorithmes, qui servent ensuite à essayer d'envoyer la bonne information à la bonne personne au bon moment.

- la géolocalisation via une adresse physique (par défaut l'adresse IP de connexion). Cela permet alors de coupler cette information avec des données de socio-économie des territoires, qui existent en dehors d'internet.

Tous ces efforts visent la rationalisation de la diffusion d'informations, voire celle de la rédaction de ces mêmes dernières selon que l'on soit homme/femme, enfant/adolescent/adulte/personne âgée, etc. Cette taylorisation, qui touche pour le moment les annonces publicitaires, peut-elle toucher d'autres types d'informations produites dans d'autres champs sociaux, notamment les contenus journalistiques ? L'ensemble des industries culturelles se développant sur Internet seront-elles concernées par cette tendance : presse écrite, télévision, cinéma, radio, édition ? Passera-t-on d'un espace public fragmenté (segmenté) à un espace public atomisé (individualisé) ?

## 2. Enjeux de la recherche

Notre réflexion s'appuiera sur une expérience formée dans le cadre d'un contrat de recherche financé par un acteur privé, éditeur de contenus et régie publicitaire sur le web<sup>3</sup>.

Du point de vue scientifique, ce travail était l'occasion, d'une part, d'accéder à un important terrain de recherche afin de traiter de la problématique de l'usage des newsletters, entre sphère publique et sphère privée. Par usage, nous entendons les rapports de dépendance réciproque, qui se créent entre les individus et le dispositif sociotechnique « lettre électronique d'information », ainsi que les dispositions et valeurs sociales qu'il actualise chez les lecteurs.

D'autre part, au-delà des questions éthiques de récolte de données sur les individus, elle permettait de l'intérieur de questionner la notion de « personnalisation » des contenus

---

<sup>3</sup> Celui-ci exploite 11 millions d'adresses emails valides en novembre 2010, dont 4 millions actives, c'est-à-dire des internautes, derrière l'adresse, ayant cliqué au moins une fois sur des contenus qui lui ont été envoyés dans le dernier mois. Cette base de données, en évolution constante, est alimentée par deux sources différentes : divers acteurs du web et un service d'inscription à une newsletter personnalisable. Seules 20% des adresses sont renseignées (informations déclarées par les personnes). L'enjeu marchand était donc, d'abord, d'enrichir ces profils avant d'établir des règles de personnalisation. D'ailleurs, le contrat de recherche spécifiait l'élaboration de règles éditoriales de production et diffusion de contenus en rapport avec des comportements observés afin d'améliorer à la fois la lecture/consommation des informations dans les newsletters et celle des emails publicitaires.

édités et sa pertinence pour les lecteurs. Et par ricochet, la pertinence économique de la publicité individualisée pour les entreprises de presse écrite pouvait également être interrogée pour compenser l'abandon du modèle du double marché (COLLET, 2007). En effet, le modèle économique de la presse écrite sur support papier suppose une régularité dans la parution (journalière, hebdomadaire, mensuelle) mais également une acquisition du support de façon privative. Le journal est vendu à des lecteurs et le nombre de lecteurs est vendu à des annonceurs. Cette double source de revenus, à laquelle il convient d'ajouter les petites annonces, est nommée la logique du double marché<sup>4</sup>. Sur le web, les entreprises de presse ont du tenir compte de l'offre d'information gratuite proposée par les « pure Player » d'Internet (portail d'information, fournisseurs d'accès, ...), des capacités d'autofinancement limitées et des cultures professionnelles des rédacteurs habitués à produire pour le support papier plutôt que pour l'écran d'ordinateur. Dans ce contexte concurrentiel sont alors apparus de nouveaux contenus (commentaires des articles, forums, chats, ...), ont été remis en cause les modes de production de l'information (contenus sans déictiques, périodicité de la publication, intégration de productions des lecteurs, ...) et leur valorisation économique (accès gratuit à l'information financée par la pub, paiement pour des services à valeur ajoutée, ventes et/ou échanges d'information avec d'autres sites).

Pour tenter d'atteindre ces objectifs, nous avons considéré l'activité de lecture comme une activité sociale, qui se comprend dans l'interaction entre un dispositif et un individu, qui sont tout à tour des objets et des sujets socialisés : les dispositifs construisent leurs « utilisateurs » autant que les dispositifs sont construits par ces utilisateurs. La volonté de faire se rapprocher le lecteur réel du lecteur modèle (ECO, 1985) est donc un travail sans fin car le lecteur réel n'est pas un être figé mais un être social pris dans des relations de dépendance réciproque avec son environnement humain, technologique, naturel, ... (ELIAS, 1997). Et l'environnement technologique de lecture n'est pas neutre sur l'individu : il le forme en instaurant des normes économiques, ergonomiques, etc. de fait (normes acceptées, voire même imposées par moment) autant qu'il est formé par l'individu.

Cela n'est pas sans incidence sur notre travail de recherche puisque le lecteur n'est pas transcendant au dispositif et que par conséquent les règles de publication risquent non seulement de ne pas être transportables à d'autres supports mais également d'évoluer dans le temps. La personnalisation ne peut être qu'un processus constant, qui nécessite le

---

<sup>4</sup> Le modèle économique du double marché conjugue le modèle de l'édition (de livre, de musique, de film) et du flot (radio et télévision). Dans le secteur des industries culturelles, on trouve également les modèles du club (canal+), du compteur (Minitel) et du courtage (information spécialisée) (KOCERGIN, 2004). Des caractéristiques de ces différents modèles se retrouvent aujourd'hui dans des sites web d'industries culturelles.

développement d'outils dynamiques de marquage des lecteurs et de puissants serveurs de traitement de l'information.

L'approche de la question a alors tout à gagner à être sémiotique car « *le propre de la sémiotique est d'avoir en vue tous les ensembles signifiants possibles, mettant donc en jeu nécessairement non seulement le rapport signifiant vs signifié, mais aussi celui du signe à ses utilisateurs, quelque soit leur extension* » (COURTES, 2003, p. 68). Tous leurs utilisateurs, producteurs et consommateurs de signes sont ainsi concernés. Il est alors nécessaire d'envisager les signes en contexte de production et de réception ayant chacun sa propre grammaire (BOUTAUD, VERON, 2007), c'est-à-dire ses propres logiques de fonctionnement (acteurs, savoir et savoir-faire) et ses propres enjeux socio-économiques.

C'est donc dans ce cadre sémiotique, que nous avons en 2008 analysé le « design » des principales newsletters existantes sur le marché français en prenant en compte différentes dimensions : le niveau spatial et plastique de la page/écran, le niveau linguistique des contenus (thématique, séquences linguistiques et énonciation), et celui des dispositions sociales<sup>5</sup> auxquelles ces informations pourraient participer : gagner de l'argent, réduire ses dépenses, jouer avec les autres, découvrir un objet/service, gérer le temps, découvrir l'expérience des autres.

Néanmoins, cette première étape de recherche nous a permis d'élaborer une offre de 3 newsletters, présupposant 3 grands comportements modèles, qui segmentent l'audience dès le départ puisqu'elles mettent en avant des lecteurs modèles différenciés.

Ces différents styles sont proposés à l'internaute qui s'abonne gratuitement sur le site dédié à cela ou qui la reçoit sans abonnement via les adresses électroniques collectées par les partenaires de l'entreprise :

---

<sup>5</sup> Cette dernière dimension n'est pas simple à déterminer car souvent elle dépend du cadre institutionnel dans lequel elle se situe. Par exemple, pour des annonces publicitaires ou des comparatifs de produits, mais également pour les messages en provenance d'ebay, il s'agira a priori de réduire ses dépenses. Mais pour un article comparant des produits, s'agira-t-il de réduire ses dépenses, de réduire son temps (de recherche d'informations), de découvrir l'expérience des autres, de découvrir un objet/service ? Tous ces critères en même temps ? Si tous les critères agissent en même temps, la qualification du contenu n'est alors plus discriminante et ne peut plus agir sur la personnalisation de l'information.

- une lettre d'information au format « petites annonces », dont la caractéristique principale est le découpage thématique des offres et une présentation d'articles les uns derrière les autres sur une seule colonne. Le comportement valorisé est a priori celui des chasseurs de bonnes affaires sans question de genre ou d'âge. C'est le modèle des sites leboncoin.fr et ebay.fr appliqué au monde des newsletters à la différence que le mode mosaïque proposé par ce dernier n'a pas été retenu.



Figure 1. Newsletter au format des petites annonces



- un format de type presse écrite, visant un public de 30 ans, ayant une appétence pour l'actualité. Le fond est blanc, les titres en bleue et les articles en noir pour mieux imiter les identités visuelles du journal LeMonde.fr et LeFigaro.fr. Les principaux principes de la mise en forme de l'information journalistique ont été respectés même si l'information et plutôt de l'information service : hiérarchisation de l'information (titres en gros caractères et en « gras », ajouts de photographies pour les titres les plus importants), colonage pour la mise en page, etc.



Figure 2. Newsletter au format de la PQR ou PQN

- une newsletter, dont le public cible est plus particulièrement une lectrice d'informations services, de plus de 20 ans, ce qui n'empêche pas que les lecteurs réels puissent être des hommes. Ce format comprend des couleurs vives. Les photographies privilégient des femmes en situation active.



**Figure 3.** Newsletter au format de la presse magazine féminine

Ces formats ont évolué dans le temps, notamment dans leurs qualités graphiques ou

encore dans leurs contenus sémantiques, mais les principes énoncés ci-après ont été gardés tout au long des différentes versions adoptées. De plus, ces versions ont pour constante d'éviter le genre dépliant commercial ; genre adopté par le marchand en ligne [cdiscout.fr](http://cdiscout.fr) pour sa lettre d'informations électronique.

### 3. Usage des newsletters

Dans le cadre de recherche qui est le notre, l'acceptation ou le refus de lire la lettre d'informations peut s'apparenter à de la fidélisation. Pour autant, la question théorique de l'usage des newsletters ne peut se limiter aux stratégies des offreurs. Pour les mêmes raisons, nous ne nous sommes pas limités à la problématique marketing de la « perception et du comportement du consommateur face à la pression commerciale », pour qui faire un effort d'ouverture puis un autre effort cognitif de compréhension du message pour ensuite ne pas cliquer génère une déception (MICHEAUX, 1997). Nous avons donc mené un travail d'analyse de l'usage de ces newsletters sur un corpus constitué à partir des adresses électroniques collectées par la société partenaire. A partir de là, nous avons cherché à élaborer des règles de publication, pour améliorer le nombre d'emails publicitaires ouverts et d'articles lus.

Le contexte de recherche nous conduit à préciser que les conclusions, auxquelles nous arrivons, ne peuvent pas être élargies à l'ensemble des internautes français. En effet, le corpus analysé est uniquement représentatif de la base d'adresses en possession de l'entreprise. Par ailleurs, les analyses se sont basées sur l'action de cliquer (ouvrir le mail reçu, lire la suite de l'article), rendant impossible de saisir l'activité sociale de lecture dans sa complexité. Des entretiens semi-directifs avec des lecteurs recrutés dans la base de données de la société partenaires sont d'ailleurs prévus afin d'enrichir nos connaissances. Néanmoins, malgré ces limites, nous avons cherché à donner une interprétation de la réactivité (taux de clic par rapport au nombre d'envois) de l'ouverture des emails et de clic sur les articles en croisant des « saillances » avec des « prégnances » selon deux approches complémentaires (PIGNIER, DROUILLAT, 2004).

La première approche visait à identifier des comportements modèles vis-à-vis de la réception de message électronique. Nous avons pour cela profité de la phase de recrutement d'internautes, par la société partenaire, qui nous a mis à disposition un échantillon de 5000 personnes. Celui-ci était composé à 51,28% d'hommes et 48,72% de femmes, nés entre 1951 et 1979. La seconde approche consistait à identifier des recoupements entre critères retenus pour qualifier les contenus et les internautes. Pour ce faire, chaque semaine, nous avons ainsi pu analyser les données sur l'ouverture de la newsletter.

Dans les deux cas, notre méthode a consisté à suivre l'évolution dans le temps des taux d'ouverture des mails et des articles et, sur cette base, à réaliser des croisements simples entre données. Enfin, l'unité de mesure favorisée a été le taux de réactivité, c'est-à-dire le taux de clic par rapport au nombre de fois où l'article a été envoyé. Notre démarche est donc plutôt qualitative que statistique.

### ***3.1. Acceptation et refus***

Pour tenter de répondre à nos interrogations sur l'usage, nous avons donc commencé par analyser les traces de consultation de 5000 individus sélectionnés parce qu'ils avaient cliqué sur au moins un message. Après plusieurs mois, l'échantillon n'était plus que de 1843 personnes, après élimination des personnes ne cliquant plus ou s'étant désabonnées. Ce suivi diachronique sur plusieurs mois nous a permis d'établir des profils de lecture, que nous avons regroupé en trois catégories : comportement de refus, comportement de lecture occasionnelle, comportement de lecture régulière. Ensuite, au sein de chacun de ces groupes, on a pu repérer des sous-groupes.

#### ***3.1.1. Les refus***

Tout d'abord, il existe des comportements de refus du service par les internautes, qui reçoivent dans leurs emails un des trois formats sans l'avoir demandé. Sont donc exclus de cette première catégorie, les lecteurs qui se sont abonnés, ce qui n'est pas vrai pour les comportements de lecture occasionnelle et régulière.

Parmi les individus, qui refusent le service, il est possible de distinguer deux sous-catégories :

- les lecteurs qui n'accrochent pas dès le départ. Soit, ils n'ouvrent pas le mail premier qu'il leur est envoyé, soit, ils se font une idée négative en une seule lecture puisque les mails suivants ne sont pas ouverts. Sans entretien semi-directif permettant d'analyser les pratiques des refusants, il est difficile de savoir si les newsletters sont en général rejetées (non usage) ou s'il s'agit d'un refus uniquement pour cette offre.

- les lecteurs découvreurs, qui se donnent plusieurs ouvertures de lettre avant de se faire une idée définitive mais qui décrochent au bout de quelques semaines. Le décrochage est alors souvent progressif : de plusieurs articles ouverts dans les premiers envois, ils passent à des ouvertures épisodiques, puis à plus aucune ouverture. Ces lecteurs sont a priori différents des précédents car le refus ne peut plus être hypothétiquement lié à un non usage des newsletters mais plutôt à un jugement négatif basé sur l'expérience.

### ***3.1.2. Les lectures occasionnelles***

Il y a ensuite les comportements de lecture occasionnelle, avec des lettres ouvertes quasiment toutes les semaines et des clics sur les articles de manière assez irrégulière. On peut, ici encore, faire une distinction entre trois types de comportement :

- les lectures ponctuelles, qui ne portent pas sur les mêmes rubriques. On peut supposer que c'est le contenu de l'article, qui attire. Cependant dans les informations recueillies, il est impossible de savoir qui du thème de l'article, du style ou de tout autre élément de mise en forme enclenche le clic.
- les lectures de type veille, où les clics sont rares mais se font presque toujours sur des articles de la même rubrique. Ici, c'est l'appétence pour un thème particulier qui guide le clic, comme c'est le cas pour le groupe 5 (cf. ci-dessous), intéressé par les informations « santé-beauté » et « mode et accessoire ». L'usage est ici clairement identifiable et s'apparente à de la rationalisation de la recherche d'informations ciblées.
- les lectures occasionnelles laissant supposer une certaine curiosité, où systématiquement les mails sont ouverts mais les articles ne sont pratiquement plus cliqués. Ces individus correspondent au groupe 1 du tableau qui suit. A-t-on affaire à des chasseurs de bonnes affaires attirés par les promotions, les nouvelles offres moins chères, etc. ?

### ***3.1.3. Les lectures régulières***

Il y a enfin les comportements de lecture régulière, c'est-à-dire des clics sur des articles à rythme régulier. Ce type de comportement est illustré par les petits groupes, 7 à 10. Dans cette catégorie, il faut introduire des distinctions entre :

- les lectures de tout, toutes semaines. Ces comportements sont très peu nombreux mais laissent supposer que l'offre est appropriée comme un rapport à une instance donnant du sens au monde. La lettre d'informations agirait alors comme espace de médiation.
- les lectures à habitudes, où systématiquement sont consultées les mêmes rubriques le même jour de chaque semaine. C'est une forme de rituel pour le lecteur, dont on peut se demander s'il est propre à la lecture des emails, plus largement de l'usage d'internet, voire de son rapport personnel au monde.

Au sein de cette première typologie, nous nous sommes plus particulièrement intéressés aux comportements de lecture régulière. Une analyse typologique de type K-Means a ainsi été réalisée sur la base du cumul des clics des individus, qui a abouti à une typologie en 10 groupes<sup>6</sup>.

Grp	n	%	Nombre moyen de clics par mois	Grp	n	%	Nombre moyen de clics par mois
1	1373	75,7	0,2	6	19	1,0	1,8
2	142	7,8	1,2	7	9	0,5	2,2
3	132	7,3	0,6	8	8	0,4	2,6
4	88	4,9	1,1	9	8	0,4	3,0
5	24	1,3	1,2	10	8	0,4	3,3

**Figure 4.** *Typologie K-Means*

Enfin, l'analyse des clics révèle que les groupes n'ont pas les mêmes centre d'intérêts ni le même degré de curiosité. Le tableau suivant présente le taux moyen de clics de chaque groupe en fonction du contenu rédactionnel de l'article. Il montre que ces taux varient grandement en fonction du groupe et du contenu.

<b>G1</b>	Techno (3,67), santé-beauté (0,35), maison-jardins (0,25), voyages-vols (0,20), mode et accessoires (0,19), gastronomie (0,18)
<b>G2</b>	Maisons/jardins (3,15), mode&accessoires (1,89), Santé-Beauté(1,59), Gastronomie (0,75), voyages et vols (0,51)
<b>G3</b>	maison-jardin (1,23), santé-beauté (1,23), sports-loisirs (0,85), techno (0,83), auto-moto (0,82)
<b>G4</b>	Gastronomie (3,86), Santé-Beauté (1,52), Techno (1), Maison-Jardin

<sup>6</sup> Trois remarques peuvent être faites concernant ce tableau. En premier lieu, la taille des groupes est très disproportionnée, avec un groupe dominant représentant 76% de l'échantillon, 3 groupes représentant entre 5 et 8% et 6 petits groupes à moins de 1,5%. En second lieu, il apparaît que le nombre de clics mensuels moyen, tout en restant faible, augmente avec la taille des groupes. En d'autres termes, plus un groupe est gros, plus les individus de ce groupe ont tendance à ouvrir et lire les articles de la newsletter.

	(0,91), Mode&Accessoires (0,74)
<b>G5</b>	Santé-Beauté (8,79), Mode&Accessoires (3,08), Maison-Jardin (1,71), Gastronomie (1,42), Vols&Voyages (1,08), Enfants (1,04), Sports&Loisirs (0,92)
<b>G6</b>	Techno (12,13), Santé-Beauté (5,89), Maison-Jardin (5,00), Gastronomie (3,63), Ecologie-Environnement (2,58), Enfants (1,84), Mode&Accessoires (1,79), Vols&Voyages (1,63)
<b>G7</b>	Finance-Assurance (7,44), Techno (3,40), Maison-Jardin (3,00), Sports&Loisirs (2,44), Santé-Beauté (2,33), Auto-Moto (1,67), Mode et Accessoire (1,63)
<b>G8</b>	Mode et Accessoire (10), Gastronomie (9,38), Santé-Beauté (6,88), Maison-Jardin (5,75), Vols et Voyages (2,00)
<b>G9</b>	Santé-Beauté (5,75), Maison-Jardin (3,88), Sports et Loisirs (2,75), Techno (2,50), Droit (2)
<b>G10</b>	Maison-Jardin (10), Gastronomie (6,63), Santé-Beauté (4,50), Mode et Accessoire (3,25), Techno (2,47)

**Figure 5.** *Typologie et thèmes préférés*

Ces groupes sont en général assez mixtes sauf le G7, fortement masculin et le G10 fortement féminin. Ce dernier est lecteur du format « magazine » alors que le premier lit plutôt les deux premiers formats. Les groupes 7 à 10 sont composés de lecteurs réguliers, contrairement aux autres groupes, qui n'ouvrent pas ou ne lisent pas la lettre d'informations toutes les semaines. Leurs appétences informationnelles se concentrent sur des thèmes. Ceci n'est pas le cas du groupe 3, qui a des goûts assez éclectiques. Les individus qui le composent ne sont pas des veilleurs mais plutôt des curieux. Au contraire, le groupe 5 regroupe des comportements de veille informationnelle : lecture irrégulière d'informations appartenant aux mêmes thèmes.

### 3.2. Genre et lecture

Il nous a semblé important de questionner le genre de l'abonné à la newsletter, parce que le secteur de la presse écrite, magazine notamment, a déjà fortement structuré son offre sur cette dimension et a su créer des habitudes chez les lecteurs. Cette dimension

structure plus fortement les usages que la dimension générationnelle (âge des lecteurs). Nos données<sup>7</sup> permettent ainsi de voir les relations suivantes :

- plus un « lecteur modèle » est marqué dans une newsletter et envoyé aux bonnes personnes, meilleur est la réactivité aux envois. Par exemple, le format, orienté « femme » et envoyé essentiellement à des femmes (déclaration du genre lors de l'inscription ou analyse du prénom dans l'adresse), est celui qui a le meilleur taux de réactivité. Les autres formats attirent certes des lecteurs mais le taux de réactivité est moins bon (de un à plusieurs points selon les semaines) comme le montre le tableau suivant, pour la semaine 52 de l'année 2010. Notons que les deux autres taux de réactivité ne sont pas statistiquement différents.

Format	Volume abouti	Ouvreurs	Tx ouv	Cliqueurs	Tx de clic	Tx de réact.
announces	801103	42199	5,27%	1889	0,24%	4,48%
femme	789780	41880	5,30%	2325	0,29%	5,55%
newspaper	868226	50358	5,80%	2141	0,25%	4,25%

**Figure 6.** Données issues de la semaine 52 de l'année 2010

- le genre permet également de comprendre l'appétence pour certains contenus. Par exemple, la rubrique auto/moto attire plus les hommes que les femmes et inversement pour la rubrique « santé et beauté », comme le montre le tableau suivant pour cette dernière rubrique :

civilité	Age	Volume	Ouvreurs	Tx ouv	Cliqueurs	Tx clic	Tx react
<b>Mr</b>	<20	91 832	8 358	9,10%	70	0,08%	<b>0,84%</b>
	20-29	97 166	4 637	4,77%	16	0,02%	<b>0,35%</b>
	30-39	127 964	10 046	7,85%	36	0,03%	<b>0,36%</b>
	40-49	107 356	11 005	10,25%	41	0,04%	<b>0,37%</b>

<sup>7</sup> Pour les besoins de l'argumentation et la bonne visualisation des données, nous avons focalisé notre attention sur les semaines 26 et 52, qui n'ont rien de particulier si ce n'est de permettre une vue sur le milieu et la fin de l'année 2010.



	50-59	53 203	6 117	11,50%	32	0,06%	<b>0,52%</b>
	60-69	16 935	2 309	13,63%	25	0,15%	<b>1,08%</b>
	70-80	2 422	325	13,42%	5	0,21%	<b>1,54%</b>
	>80	846	106	12,53%	4	0,47%	<b>3,77%</b>
<b>Total Mr</b>		<b>497 724</b>	<b>42 903</b>	<b>8,62%</b>	<b>229</b>	<b>0,05%</b>	<b>0,53%</b>
<b>Mrs</b>	<20	40 531	3 075	7,59%	48	0,12%	<b>1,56%</b>
	20-29	7 486	181	2,42%	0	0,00%	<b>0,00%</b>
	30-39	11 561	685	5,93%	5	0,04%	<b>0,73%</b>
	40-49	12 412	842	6,78%	9	0,07%	<b>1,07%</b>
	50-59	9 207	670	7,28%	9	0,10%	<b>1,34%</b>
	60-69	4 047	380	9,39%	7	0,17%	<b>1,84%</b>
	70-80	623	57	9,15%	2	0,32%	<b>3,51%</b>
	>80	115	11	9,57%	0	0,00%	<b>0,00%</b>
<b>Total Mrs</b>		<b>85 982</b>	<b>5 901</b>	<b>6,86%</b>	<b>80</b>	<b>0,09%</b>	<b>1,36%</b>

**Figure 7.** Données issues de la semaine 26 de l'année 2010

- en ce qui concerne le genre, le tableau ci-dessus permet également de noter une différence notable en termes d'ouverture des newsletters et de lecture des articles. En effet, si le taux d'ouverture des newsletters est meilleur chez les hommes que chez les femmes (8,62% contre 6,86%,  $p < 0,000$ ), la réactivité est meilleure chez les femmes que chez les hommes (1,36% contre 0,53%). Ce résultat suggère que les femmes font peut-être plus confiance au nom de l'expéditeur et/ou au titre du mail pour décider même si la différence n'est pas statistiquement significative.

### 3.3. Formats et lecture

A l'inverse, quel que soit le format, les mêmes zones sont les plus lues à l'écran. A ce niveau, il n'y a pas de différence entre les sexes et les âges. C'est plutôt une acculturation à lecture haut/bas et gauche/droite, qui explique ce phénomène. En effet, les articles les plus cliqués sont positionnés en haut de l'écran sur le côté gauche. Et plus on descend dans la page, moins les articles sont cliqués, comme le montre les tableaux ci-dessous.

« Femme » et « Petites annonces »			Succès des zones pour « Femmes			
Edito (Z1)			zone	cliqueurs uniques	clics	Taux réact
Image	Titre		5	1033	1599	0,19%
	Développé	Lire la suite	6	407	506	0,07%
Texte (Z5)			14	228	276	0,04%
Image	Titre		7	210	261	0,04%
	Développé	Lire la suite	13	193	264	0,03%
Publi (Z2)	Rubrique		16	193	240	0,03%
Image	Titre		8	184	223	0,03%
	Développé	Lire la suite	9	175	209	0,03%
Texte (Z6)	Rubrique		10	161	201	0,03%
Image	Titre		11	160	197	0,03%
	Développé	Lire la suite	12	122	135	0,02%
Texte (Z7)	Rubrique		15	118	136	0,02%
Image	Titre		17	115	145	0,02%
	Développé	Lire la suite	2	61	66	0,01%
Publi (Z3)	Rubrique		4	50	56	0,01%
Image	Titre		3	42	47	0,01%
	Développé	Lire la suite				
Texte (Z8)	Rubrique					
Image	Titre					
	Développé	Lire la suite				
Texte (Z9)	Rubrique					
Image	Titre					
	Développé	Lire la suite				
Publi (Z4)	Rubrique					
Image	Titre					
	Développé	Lire la suite				
Texte (Z10)	Rubrique					
Image	Titre					
	Développé	Lire la suite				
Surprise (Z11)	Rubrique					
Image	Titre					
	Développé	Lire la suite				
Surprise (Z12)	Rubrique					
Image	Titre					
	Développé	Lire la suite				
Desinscription						
Parrainage						
Liste des rubrique						

  

Succès des zones pour « Petites annonces »			
zone	cliqueurs uniques	clics	Taux réact
5	803	1193	0,21%
6	362	439	0,10%
7	194	280	0,05%
9	151	204	0,04%
8	146	212	0,04%
10	120	164	0,03%
2	78	88	0,02%
16	51	63	0,01%
3	50	61	0,01%
13	46	71	0,01%
4	43	45	0,01%
11	40	48	0,01%
14	28	32	0,01%
17	26	27	0,01%
12	24	29	0,01%
15	24	28	0,01%

**Figure 8.** Découpage et succès des zones pour « femmes » et « petites annonces »

Edito (Z1)			cliqueurs	Taux
zone	uniques	clics	réact	
5	1150	1696	0,18%	
6	422	535	0,07%	
13	296	372	0,05%	
7	233	288	0,04%	
8	189	240	0,03%	
14	186	230	0,03%	
16	178	219	0,03%	
9	174	220	0,03%	
12	136	169	0,02%	
10	125	162	0,02%	
17	112	156	0,02%	
11	105	148	0,02%	
2	91	99	0,01%	
3	77	90	0,01%	
15	75	91	0,01%	
4	59	68	0,01%	

  

Image	Titre	
	Développé	Lire la suite
Texte (Z5)		
Image	Titre	
	Développé	Lire la suite
Titre Texte (Z6)	Titre Texte (Z7)	Titre Texte (Z8)
Image	Image	Image
Publi (Z2)		
Développé		Lire la suite
Publi (Z3)		
Développé		Lire la suite
Publi (Z4)		
Développé		Lire la suite
Texte (Z9)	Rubrique	
Image	Titre	
	Développé	Lire la suite
Texte (Z10)	Rubrique	
Image	Titre	
	Développé	Lire la suite
Surprise (Z11)	Rubrique	
Image	Titre	
	Développé	Lire la suite
Surprise (Z12)	Rubrique	
Image	Titre	
	Développé	Lire la suite

**Figure 9.** Découpage et succès des zones pour « News »

Il est alors possible de supposer que les lecteurs, habitués aux « arts de faire » de la presse en général, qui hiérarchise l'information dans la hauteur des pages et la profondeur des sites, soient attirés par les premiers articles auxquels ils attribuent une importance supérieure. Il est aussi normal que l'édito et les publicités soient délaissés. Mais comme la newsletter est reçue dans le courrier électronique et n'apparaît pas au même endroit pour tous les lecteurs selon le logiciel ou le service de webmail, il est hasardeux de formuler des conclusions sur la mise en page et la lecture.

### 3.4. Rapport information au temps

Le fait marquant est la relation entre la date de l'envoi et la date de la lecture, qui peut s'étaler sur plusieurs semaines, comme le montre le tableau qui suit :

<b>date envoi 07 juin 2010</b>	<b>date ouverture</b>	<b>Somme de Tx ouv</b>	<b>Somme de Tx reac</b>
	<b>L 07-juin</b>	100,00%	5,00%
	<b>M 08-juin</b>	100,00%	5,42%
	<b>Me 09-juin</b>	100,00%	6,23%
	<b>J 10-juin</b>	100,00%	7,55%
	<b>V 11-juin</b>	100,00%	6,40%
	<b>S 12-juin</b>	100,00%	5,21%
	<b>D 13-juin</b>	100,00%	4,86%
	<b>L 14-juin</b>	100,00%	4,89%
	<b>M 15-juin</b>	100,00%	4,95%
	<b>Me 16-juin</b>	100,00%	5,51%
	<b>Je 17-juin</b>	100,00%	2,46%
	<b>V 18-juin</b>	100,00%	6,33%
	<b>S 19-juin</b>	100,00%	5,88%
	<b>D 20-juin</b>	100,00%	6,67%
	<b>L 21-juin</b>	100,00%	2,90%

**Figure 10.** *Ouvertures au fil des jours pour une même newsletter*

Ce constat ne recoupe pas la plupart des études en la matière, dont celle menée par le Syndicat National de la Communication Directe (SNCD). La raison se trouve du côté du contenu et notamment du succès des marronniers, à savoir des informations revenant chaque année à la même période : la rentrée des classes en Septembre, les cadeaux à Noël, etc. En effet, on peut constater dans nos analyses que l'information en lien avec une période donnée a de meilleur taux de réactivité que l'information intemporelle. Doit-on une nouvelle fois y voir une habitude construite par les entreprises de presse écrite et les médias d'information en général ?

Quant aux productions lues mais n'ayant pas de rapport avec la saison, les titres formulant une promesse et/ interpellant tendent à attirer les clics plus que les autres. Ainsi, dans la liste des titres mentionnés ci-dessus, on peut repérer dans le cadre d'un travail en cours :

- les titres promettant de réaliser les choses par soi-même : « Réalisez vos propres meubles en carton », « Comment organiser un vide-grenier », « Comment payer moins d'impôts »,

- les titres mettant en cause les compétences du lecteur ou proposant une aide : « Comment bien choisir son nouveau vélo ? », « Savez-vous gérer votre temps ? », « Etudes de commerce, que choisir ? »

#### 4. Conclusion

Les organisations de production de contenus médiatiques sont des configurations sociotechniques complexes à l'inverse de la réception qui n'est pas organisée et individuelle (BOUTAUD, VERON, 2007). Pourtant, les organisations de production privées travaillent à partir de chiffres et de données sur des regroupements de consommateurs pour mieux cibler leurs offres et maximiser les profits. Les constructions ainsi fabriquées sont des idéaux-types statistiques permettant de segmenter la consommation en répondant à deux questions complémentaires : qui consomme quoi et quoi est consommé par qui ? Cette manière de penser la segmentation est valable pour les médias de masse. Qu'en serait-il d'un média qui ne serait pas de masse ? La question se pose pour l'ensemble des dispositifs de personnalisation de l'information sur Internet.

Deux grandes conclusions ressortent de cette recherche, toujours en cours. La première porte sur les limites des stratégies de personnalisation de l'information. La seconde sur l'usage des lettres d'information électronique diffusées par mail.

Même avec des attributs permettant des différenciations fines, on se rend compte que de nombreux lecteurs de même âge et/ou de même sexe sont attirés par le même type d'information, et adoptent les mêmes comportements de lecture. En d'autres termes, l'activité individuelle de lecture est fortement déterminée socialement. C'est nettement visible dans l'analyse des zones les plus lues, où les habitudes formées par la presse sur support papier conduisent le lecteur à privilégier les informations en haut de la page comme les ont déjà montré les travaux de Moullaud et Tétu (1996). Mais ce l'est aussi dans les appétences genrées pour certains contenus. Les lecteurs ne recherchent donc pas seulement des informations mais aussi des espaces d'interprétation du monde orienté en fonction d'une visée, assumé par une instance, intégré dans un dialogue constant avec ses lecteur réels ou imaginaire (TETU, 1999 ; ESQUENAZI, 2002).

En ce qui concerne les usages, plusieurs comportements de lecture (comportements de veille, ouverture de la newsletter mais aucune ouverture d'articles, lectures réservées aux mêmes jours et plus ou moins mêmes heures de la semaine, ...) et le succès relatif de titres portant sur les compétences situent les lettres d'informations étudiées dans des usages rationnels de ces supports par rapport au temps. Ils peuvent donner l'illusion ou réellement permettre au lecteur du gagner du temps : comportements de veille informationnelle, augmentation de son capital de connaissances, etc. Tout comme pour

la lecture induite par la presse écrite sur support papier, c'est le cadre de la lecture des emails, qui pourrait fortement structurer cet usage. En effet, l'étude des usages des premières messageries sur minitel montraient déjà que la boîte aux lettres est un outil de communication périphérique au téléphone (JOUET, TOUSSAINT, 1991). Elle a une fonction de répondeur, qui permet de rester en contact avec des interlocuteurs tout en gérant mieux son temps ou en permettant d'organiser des pratiques de groupe (se fixer des RDV pour aller au cinéma, etc.). Elle se distingue des autres dispositifs informatiques de communication synchrone (chat) ou asynchrones (forums, blogs).

Pour conclure, il est légitime de se demander si les stratégies de personnalisation de l'information, au lieu d'emprunter des voies menant à une extrême individualisation du rapport à l'information, ne doivent pas plutôt favoriser une segmentation du lectorat en petits groupes tous marqués par le sceau de la rationalité par rapport au temps. Dès lors, les acteurs du web devraient plutôt opter pour des « individualisations massives » de leurs contenus. Ces dernières seraient d'autant plus justifiées que, comme le démontre les succès des groupes d'amis ou de fans sur Facebook et Twitter, on est plus dans du marketing « one to few » que « one to one » en terme de personnalisation.

## Bibliographie

- BERNERS-LEE, T., HENDLER J., LASSILA O., *The Semantic Web*, Scientific American, May 2001,
- BOUTAUD JJ., VERON E., *Sémiotique ouverte. Itinéraires, sémiotiques en communication*, Lavoisier, 2007, 194 p
- BUSH Vannevar, As We May Think , Atlantic Monthly, juillet 1945
- CARON ML, FRANCONY JM, QUINETTE N, « intégrer différentes propositions de valeur pour renforcer son business model » in *Business models dans l'innovation* (sous la direction de V. Chanal. PUG 2011)
- COLLET L., De l'édition de presse papier à l'édition de presse sur Internet : émergence d'un nouveau bien culturel et hybridation des modèles socio-économiques de production, in Actes du colloque *on line services, networks, contents, usages*, Paris : France, 2007
- COURTES J., *La sémiotique du langage*, Nathan VUEF, 2003, 127 p.
- ECO U., *Lector in fabula ou la coopération interprétative dans les textes narratifs*, tr. k. de Myriem Bouzaher, Paris, Grasset, 1985
- ELIAS Norbert, *La société des individus*, Agora, Pocket,, Avril 1997, 301 p.
- ESQUENAZI JP, *L'écriture de l'actualité. Pour une sociologie du discours médiatique*, Grenoble : PUG, coll. La communication en plus, 2002. Cit. p6-8.

- HABERMAS J., *L'Espace Public. Archéologie de la Publicité comme dimension constitutive de la société bourgeoise (1962)*, trad. française de Marc B. de Launay, éditions Payot, Paris, 1978, 324 p.
- JOUET J, TOUSSAINT Y, *La télématique interpersonnelle, étude du courrier électronique privé*, Rapport CNET/IRIS, Juin 1991
- KOCERGIN S., *Internet à la recherche d'un modèle socio-économique. Portail : une perspective viable ?*, Thèse de doctorat de 3<sup>ème</sup> cycle, Volume 1, Décembre 2004, 280 p.
- MICHEAUX A., *Marketing de bases de données*, Editions d'Organisation, deuxième édition, Octobre 1997, 282 pages.
- MOUILLAUD, TETU, *Le Journal quotidien*, Broché, Pul, Octobre 1996
- OUAKRAT A. et ali., Les régies publicitaires de la presse en ligne, page 133 à 161 in Réseaux n° 160-161, *Presse en ligne*, 2010/2-3, p. 133-161
- PIGNIER N., DROUILLAT B., *Penser le webdesign. Modèles sémiotiques pour les projets multimédias*, L'Harmattan, Paris, 2004
- SNCD, *Étude Email Marketing Attitude 2009*, publié en 2010 et téléchargeable sur <http://www.sncd.org>
- TETU JF, « Introduction » In Jamet Claude et Jannet Anne-Marie, *La mise en scène de l'information*, Paris/Montréal : L'Harmattan, coll. Champs visuels, 1999, p 1-13.